

スピーキングテストにおいて
テスト形式の違いが受験者の発話に与える影響
ーインタビューテストとグループ・オーラルテストを中心にー

曲 明

松 山 大 学
言語文化研究 第30巻第1号（抜刷）
2010年9月

Matsuyama University
Studies in Language and Literature
Vol. 30 No. 1 September 2010

スピーキングテストにおいて テスト形式の違いが受験者の発話に与える影響 ーインタビューテストとグループ・オーラルテストを中心にー

曲 明

概 要

本研究では、学習者同士の話し合いを評価するグループ・オーラルテストの有効性を検証するため、従来のスピーキングテストとして使用されてきたインタビューテストとの比較研究を行った。実験では、高校3年生54人が両テストを受け、テスト形式の違いが受験者の発話に流暢さ、正確さ、複雑さ、発話量の観点で及ぼす影響を分析した。その結果、テスト形式の違いが流暢さと発話量へ及ぼす影響が有意となり、正確さ、複雑さに影響しないことが明らかになった。グループテストをスピーキング能力の評価に用いることによって、時間的なコスト削減ができるが、しかし、テストとしての妥当性と公平性を高めるには、発話量の統制や発話の自由度を限定できるタスクの開発などの課題も残っていることが示唆された。

キーワード：スピーキングテスト、グループ・オーラルテスト、インタビューテスト、テストパフォーマンス

1. は じ め に

近年、外国語教育の現場では学生のスピーキング能力の育成およびその学習支援が急務となっている。テストの波及効果を考えれば、スピーキング能力の育成には、テストを行うことが効果的である。しかし、一般教育現場の大規模

教室における現状として、スピーキングテストを従来のインタビュー形式で行う場合、教師と学生一対一では、時間がかかりすぎるという困難な側面がある。そこで、本研究では外国語教育におけるスピーキング能力の育成に向けた、実践的かつ妥当なテスト形式の検証と提案を目標とする。とりわけ、学習者同士の話し合いを評価するグループ・オーラルテスト（以下グループテストと略す）とインタビューテストと比較をすることによって、テスト形式の違いが受験者の発話に与える影響を明らかにしたいと考える。

2. 先行研究と本研究の目的

2.1 インタビューテスト

現在行われている直接スピーキングテスト（direct speaking test）の形式は、大きく分けると、対面形式を採る「人対人」型のテストと、録音機材に受験者が発話を吹き込む、いわば「人対機械」型のテストがある（Luoma, 2004）。インタビューテストは、「人対人」タイプのテストとして、最も一般的に行われている方法である。

Young (1992) は、インタビューテストを次のように定義している。「言語テストとしてのインタビューテストは、一方が熟達者（母語話者か、準母語話者）で、もう一方が非母語話者か、学習者であるような、対面式の話し言葉による相互行為活動である（中略）。熟達者にとってのインタビューテストの目的は、非母語話者の話す能力を評価することである。」上述概念からもわかるように、常に会話の主導権の有無によりインタビュアーと受験者の間に不均衡な力関係が存在する。インタビュアーと受験者の関係が非対称的であるがゆえに、多様な現実のコミュニケーション場面の一部しか表せないといえる（Lazaraton, 1992; Young & Milanovic, 1992）。従って、インタビューを用いて測られる受験者の会話能力は実際の能力の一部分に過ぎない。このため、参加者の関係が対等な場面での会話能力や、より総合的な会話能力の評価を目指すには、インタビュー以外のテストの利用が必要となる。

2.2 グループテストについて

インタビューテストと比較して、グループテストは学習者同士の話し合いを評価する形式のテストであるため、会話参加者間の力関係が対等になり、日常的なコミュニケーションに近い会話を引き出せるといわれている (Swain, 2001; Fulcher, 2003 など)。

今まで報告されたグループテストの中で、とりわけ利害関係の高いグループテスト (high-stakes tests) には、以下のような事例が挙げられる (表 1)。

また、表 1 にまとめられたグループテストには大まかな共通点として、次のような特徴がみられる。

- ・グループの人数：3～5 人。
- ・トピック：すべての受験者にとって、特定の背景知識、性別等で不公平の生じない話題。

表 1 利害関係の高いグループテスト

著 者(年)	テスト名称	人 数	時 間(分)	評価方法と観点
Morrison & Lee (1985)	The Hong Kong Advanced/Supplementary Level Examination	3	準備 (2) 会話 (10)	5 段階評価 ・語彙と文法 ・発音 ・流暢さ ・会話管理
Kobayashi, Johnson & Van Moere, (2005); Van Moere (2006)	The Kanda English Proficiency Test	3～4	準備 (1) 会話 (6～7)	4 段階評価 ・発音 ・語彙 ・文法 ・流暢さ ・コミュニケーション能力
Nagasawa & Tanabe (2001)	The interactive English Forum	3～4	準備 (2) 会話 (10)	10 段階評価 ・伝達性 ・協調性 ・適確さ
He & Dai (2006)	College English Test-Speaking English Test	3～5	準備 (1) 会話 (4.5)	5 段階評価 ・正確さ ・語彙 ・会話管理能力 ・会話の適切さ

- ・時間：準備に1～2分程度が与えられ、話し合う時間は5～10分前後。
- ・話し合いの進行：受験者のみで行う。但し、発話者に偏りが見られたり、会話が停滞した場合には、試験官が何らかの介入を行う。

先述の「会話参加者間の力関係が対等である」と言及した利点以外にも、グループテストには受験者一人当たりにかかる時間が少なく、効率的である(Hilsdon, 1991; Bonk & Ockey, 2003), また、評価者は質問や会話の進行をする必要がないので、複数の受験者を同時に評価できる(Bonk & Ockey, 2003), 更に、多くの受験者自身が好む傾向にあり、受験者の不安が少ない(Fulcher, 1996) テスト形式であるといわれている。

しかし、グループテストの問題点として、受験者の発話が会話を組む相手の発話に影響を受ける可能性があることも指摘されている(Kobayashi, Johnson & Van Moere, 2005; Van Moere, 2006)。受験者を相互行為の中で同時に評価できることはグループテストの利点であるが、対話者やグループ内の相互作用が各受験者の発話や評価結果に影響を与えることはグループテストにおいて避けられない問題である。従って、グループテストの利点を生かしながらも、受験者間の相互作用に対する理解を深めていく必要があると考えられる。

2.3 スピーキングテストのモデル

McNamara (1996) によると、図1のようなスピーキングテストにおける受験者パフォーマンスのモデルが提唱されている。モデルの中心には受験者のパフォーマンスを位置づけ、このパフォーマンスに影響を与える要因として、対話者、受験者、タスクの3つを挙げている。特に、受験者と対話者の間には、双方向の矢印で示され、両者の相互作用がパフォーマンスに影響することを説明している。こうした事情から、対話者もまたテストパフォーマンスに影響を及ぼす可能性を抱えており、見過ごすことはできない要因のひとつである。本研究で比較するインタビューテストとグループテストの大きな違いは、上述のモデルにおける「対話者」にある。先に述べたとおり、インタビューテストは

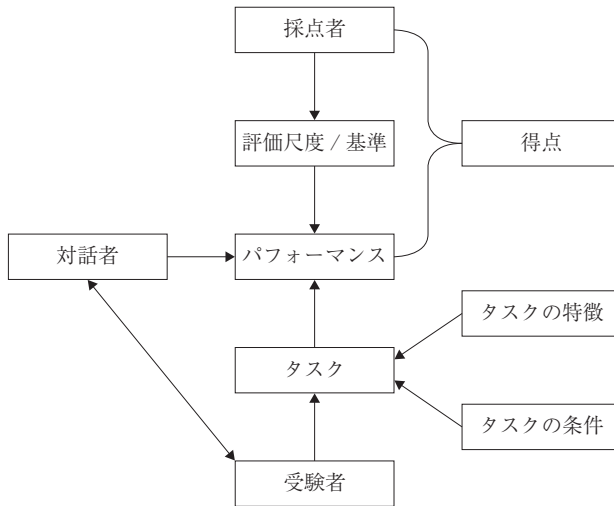


図1 スピーキングテストにおけるパフォーマンスモデル
(McNamara, 1996 より翻訳)

母語話者または母語話者に近い言語熟達者である試験官と受験者との一対一の会話であるに対して、グループテストは受験者同士が会話する。この「対話者」の違いからテストパフォーマンスに何かの影響を与える可能性があるとするれば、その違いを量的、及び質的にも精査していく必要があるだろう。

2.4 テスト形式の違いによる発話パフォーマンスへの影響

スピーキングテストで広く用いられている分析法は、評価尺度を用いたものと、ユニットごとに区切って測定する談話分析的手法が挙げられている (Ellis, 2003)。まず評価尺度を使って、テスト得点の比較結果を報告したものは、Fulcher (1996), Shohamy, Reves, & Bejarano (1986) がある。両者の研究に関して、いずれもグループテストは他の形式のテストと得点の相関が低い結果を示している。しかし、テスト得点には、採点手続きの影響を含んだものである。対話者の影響に焦点を当て、テストの妥当性を検証するには、採点手続きの影響を受けない受験者の発話そのものから分析する必要がある (Van

Lier, 1989)。従って、本研究では、近年注目されているユニットごとに区切って行う談話分析の分析法を試みる。グループテストとインタビューテストでの発話を談話分析手法で分析を行った研究には堀川（2006）が挙げられる。堀川は韓国人の日本語学習者を対象にグループテストとインタビューテストを行い、発話における流暢さ、正確さ、複雑さについて分析した。結果として、正確さのみグループテストの方が有意に高いことが分かった。堀川は上述の研究結果について、「グループテストでは、受験者自身に会話のコントロールが任されているため、形式への注意もでき、誤用の起きない日本語で話すことができた」と自身の見解を報告している。

しかし、堀川の研究では日本に在住する成人の韓国人、つまり第二言語習得環境（目標言語を使用する地域で言語習得する環境）を対象に行った研究であり、外国語習得環境（自国で外国語を学ぶ環境）の学習者にまで結果を一般化するには実証が不十分と考える。また、受験者人数（6名）が少ないため、再検証の必要があるだろう。

外国語学習者の発話特徴の解明には、発話量も重要な指標であると思われる。とりわけ、外国語学習者のコミュニケーション能力を向上させるためには、積極的な発話を促す必要がある。テスト場面においても、受験者の発話量が十分に引き出せない場合、評価不能な実態が想定され、教育上の十分なフィードバック機会を逸した状況が懸念されるだろう。Taylor（2000, 2001）はケンブリッジ英語検定試験を用い、インタビューとペア形式のスピーキングテストの発話量について研究を行った。その報告書によると、ペアテストでは、インタビューテストの75%程度の発話量に減ったと報告している。上記ペアテストにおける結果を熟考した場合、本研究対象のグループテストには、会話参加者の増員が発話量にいかなる影響を及ぼすのか、注目すべき側面を含んでいるといえよう。

上述でまとめたとおり、これまでの研究には調査参加者の規模や多様性、ひいては発話データの測定・評価方法で改善すべき課題を残している。また、テ

スト形式そのものへのアプローチは非常に少ない。本研究ではグループテストとインタビューテストのテスト形式の違いによる影響を、次節にて詳述する発話の流暢さ、正確さ、複雑さ、発話量の指標に注目し、検証していきたいと考える。

2.5 本研究で用いる発話分析の指標

先に述べたとおり、本研究ではグループテストとインタビューテストでの受験者の発話の特徴を分析するため、発話の流暢さ、正確さ、複雑さ、発話量に焦点を与える。

(1) 流暢さの指標

流暢さの指標の測定方法には、大別して、会話の一時的な停止を捉えた場合と、言い淀みを捉えた場合の2つに分類される。前者については、ポーズの数を数える方法やポーズ及び沈黙の総時間を測定する方法が考案されている(Foster, 1996; Mehnert, 1998)。一方、後者については非流暢さの尺度として、繰り返し、出だしの言い間違い、自己訂正などを数える方法がある(Foster & Skehan, 1996)。本研究では、ポーズ、沈黙の使用と該当となる言語使用の両側面に注目し、以下の指標の出現頻度を求めた。

- ・ 1分間当たりの無声 (unfilled) ポーズの数
- ・ 1分間当たりの有声 (filled) ポーズ数
- ・ 1分間当たりの繰り返しの数
- ・ 1分間当たりの自己訂正の数
- ・ 1分間当たりの出だしの言い間違いの数

(2) 正確さの指標

正確さを表す指標については、大きく分けて特定の文法項目に焦点を当てる測定方法と、一般的な正確さを測定する方法の2つに区分される。前者については、ある特定の文法項目に関するエラーを測定する方法 (Ellis, 2004) があり、後者については、特定の文法項目に限定せず、すべての節に対し、文法誤

りのない節 (Error-Free Clauses: EFC) の割合を算出する方法が提案されている (Foster & Skehan, 1996)。

本研究で使用するテストは、特定文法項目の学習成果をみる主旨ではなく、会話における全般的な言語使用の正確さに関心がある、従って、後者の分析方法を採用し、具体的に以下の2つの指標から割合を算出した。

- ・誤りのない節数
- ・誤りのない AS ユニット

(3) 複雑さの指標

複雑さに関しては、大きく分けて統語的複雑さと語彙的複雑さの2つに区分される。前者の統語的複雑さについては、AS ユニット当たりの節の数の割合を測定する方法 (Foster & Skehan, 1996) とユニット当たりの語数を測定する方法 (Mehnert, 1998) がある。一方、後者の語彙的複雑さについては、名詞、形容詞、副詞、動詞の異語数 (types) の合計を測定する方法 (Foster, 1996) と総語数に対する異語数の割合を測定する方法 (Robinson, 2001) がある。

本研究では上述に揚げた測定方法の中で、適切な測定方法とされる AS ユニット当たりの節数 (Foster & Skehan, 1996) によって統語の複雑さを測定し、総語数に対する異なり語数の割合によって語彙の多様性を測定した。

(4) 発話量指標

発話量に関しては、発話時間を測定する方法と発話の語彙数を測定する方法がある。発話時間を測定する方法は話者の性格、発話のスピードなどの要素が発話量に影響を及ぼすため、本研究では、発話語彙数を測定する方法を取る。一定時間に発話されたすべての語彙数から、単語1語からなる間投詞や繰り返し、言い淀み、自己訂正した語の数を引いた数にした。

ここで、本研究で用いる流暢さ、正確さ、複雑さ、発話量の指標は表2にまとめる。

表 2 分析の指標

	略記号	指 標
流暢さ	F 1	1 分間当たり無声 (unfilled) ポーズ数
	F 2	1 分間当たり有声 (filled) ポーズ数
	F 3	1 分間当たり繰り返しの数
	F 4	1 分間当たり自己訂正の数
	F 5	1 分間当たりの出だしの言い間違いの数
正確さ	A 1	誤りのない節数
	A 2	誤りのない AS ユニット
複雑さ	C 1	1 AS ユニット当たりの節数
	C 2	1 語当たりの異なる語数 (type token ratio)
発話量	W 1	一定時間に発話された語彙数

2.6 本研究の目的

本研究では、日本の高校 3 年生である中国語学習者を対象に、グループテストとインタビューテストを実施し、両テスト間の発話に違いがあるか明らかにしたいと考える。具体的には発話の流暢さ、正確さ、複雑さ、発話量を示す指標に注目し、これらの量的分析を通して、テストの形式違いが受験者の言語パフォーマンスにどのような影響を与えるかを探る。従って、研究課題は以下のとおりである。

研究課題：グループとインタビューによるテスト形式の違いは受験者の発話に流暢さ、正確さ、複雑さ、発話量の観点でどのような影響を及ぼすのか。

3. 研 究 方 法

3.1 テストの概要

テストの位置づけ

本研究は、東京都に所在地を置く私立大学の附属高等学校に協力を頂いた。参加者は同校に在籍する 3 学年を対象にしており、科目「中国語スピーキング」の期末考查を通じて、受講生の発話データを採取した。当該校では、中国語を

第2外国語の必修科目に定めており、受験者全員が3年の間最少でも週5時間中国語を履修していた。また、当校の卒業生はほぼ全員が附属する大学に入学するため、今回の期末テストを含めた各科総合成績は希望する学部への入学許可で大きな比重を占めていた。従って、受験者たちにとって今回の試験は利害関係の大きいテストだったといえる。

受験者とテストの組み合わせ

受験者は計54名の男子生徒のみで構成されていた。テストパフォーマンスへの影響を考え、当該の受験者の中で54人中に極端に外向的あるいは内向的な生徒がいないことを確認した上で、グループテストは3人一組で実施した。なお、生徒の組み合わせは科目担当の教師が無作為に指定した。一方、インタビューテストは教員1名が生徒個別に実施した。

テストの手順

テストの順序効果を全体で最小化するために、予め受験者を2群に分割し、双方が平衡した順序でテストを受験した。つまり、27人からなるA・Bのグループに分け、Aグループの受験者はグループテストからインタビューテストの順番で受けて、Bグループはインタビューテストからグループテストの順番で受けるように指示した。

グループテストは話題の書かれているカードを提示して、準備時間1分を与えた後、話し合いの時間にはおよそ4分を設けた。テストの進行は基本的に受験者に任せたが、発話量のばらつきが顕れた場合にのみ、教師が介入することによって調整した。インタビューテストは準備時間を与えず、受験者各自におよそ2.5分の質疑を行った。尚、テストの進行はインタビュアーである担当教員によってすべて統制した。

タスクのデザインとトピックの選定

一般に、グループ・ディスカッションについて言及した場合でも、タスクのタイプは様々である。Richards (2001)によれば、グループ・ディスカッションのタイプは5種類あり、パズル (jigsaw)、情報差 (information-gap)、問題

解決 (problem solving), 意思決定 (decision-making), 意見交換 (opinion exchange) を挙げている。本研究の受験者たちの中国語学習歴は3年であり、実施校の使用教科書が初級・中級学習者向けだったものも考慮し、「情報差」のタスクと「問題解決」のタスクを選定した。

また、今回のスピーキングテストは期末考査の趣旨が前提にあるため、テストとしての「内容妥当性」を保つ必要があった。よって、両テストで用いるトピックはその学期で使っていた教科書¹⁾から選出したものである。なお、複数グループで同時にディスカッションを展開できるよう、グループテストのタスクは3種類用意した(詳細は付録1参照)。尚、インタビューテストで用いる質問文は使用教科書からそのまま抜粋した(具体例は付録2を参照)。

3.2 分析方法

録音機材により収集した54名の発話データはすべて書き起こした。その文字化データをASユニット²⁾に分けた後、受験者の発話における流暢さ、正確さ、複雑さ、発話量を表す各指標の値を算出した。そして両テストの平均値に差があるかどうかを明らかにするため、tテストを用いて検証した。

3.3 コーディングの方法及び採点者間一致率

ASユニットの数は (Foster, 2000)³⁾ の基準に従った。流暢さ、正確さ、複雑さ、発話量のコーディング方法について、まず全18グループの中から2組の発話データを取り出し、2名の協力者と協議しながらコーディングを行った。初めに定義を確認した後、別々に作業し、一致しなかった箇所については互いが納得できるまで話し合い、話し合って解決した。このようにして分類基準を確定した後、残りのデータを更に中国語教育専攻の中国語母語話者5名の協力により処理した。最終的には、筆者が全データを確認し、採点者間一致率を算出したところ、81%~92%であった。

4. 結 果

4.1 流暢さの結果

流暢さの5つの指標において、記述統計、検定結果は表3に示したとおりである。5指標のうち、「繰り返しの数」は、グループテストの平均値が高いが、一方「有声 (filled) ポーズ数」, 「無声 (unfilled) ポーズ数」, 「出だしの言い間違いの数」, 「自己訂正の数」においては、インタビューテストの平均値が高かった。以上のことから、グループテストでの発話がインタビューテストでの発話より流暢である傾向が示唆された。しかし、tテストの結果、有意差が確認された指標は、「無声ポーズ数」($p < .01$), 「有声ポーズ数」($p < .05$) の2項目のみであった。

表3 流暢さについての記述統計及びt検定の結果

	インタビュー		グループ		t検定	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>p</i>
無声ポーズ数 (unfilled)	6.31	.76	3.98	.23	-4.40**	.00
有声ポーズ数 (filled)	5.43	.57	3.67	.44	-3.49*	.04
繰り返しの数	3.8	.61	5.16	.90	-1.81	.25
自己訂正の数	3.6	.43	3.23	.39	-0.15	.79
出だしのいい間違いの数	.47	.22	.35	.37	-1.51	.24

(注) $N=54$, $df=53$, *は5%水準で有意 ($p < .05$), **は1%水準で有意 ($p < .01$)。

4.2 正確さの結果

正確さの指標に関しては、記述統計は及び検定結果は表4に示した。該当の2指標について、「誤りのない節数の割合」, 「誤りのないASユニットの割合」両方とも、グループテストでの平均値が高かった。この結果から、グループテストでの発話はインタビューテストの発話より正確である傾向が見られた。しかし、t検定を行ったところ、いずれの指標からも有意差は検出されなかった。

表4 正確さについての記述統計及びt検定の結果

	インタビュー		グループ		t検定	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>p</i>
誤りのない節数の割合	.31	.11	.34	.13	-1.27	.21
誤りのないASの割合	.21	.09	.23	.08	-1.33	.20

(注) $N=54$, $df=53$

4.3 複雑さの結果

発話の複雑さに関しては、1 AS ユニットあたりの語数により統語の複雑さ、そして総語数に占める異なり語数の割合により語彙の複雑さを調べた。2 指標とも平均値はインタビューテストの方が若干高かった(表5を参照)。このことから、インタビューテストでの発話はグループテストでの発話より若干複雑である傾向が示唆された。しかし、t 検定の結果では、2 指標とも有意差は得られなかった。

表5 複雑さについての記述統計及びt検定の結果

	インタビュー		グループ		t検定	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>p</i>
1 AS ユニット当たりの節数	1.23	.21	1.14	.34	-5.35	.61
総語数に占める異なり語数の割合	34.10	3.77	32.26	4.35	.10	.55

(注) $N=54$, $df=53$

4.4 発話量の結果

発話量について、記述統計及び検定結果は表6に示したとおりである。計算の結果、一定時間に発話された語彙数は、グループテストよりインタビューテストの方が多く、t 検定を行ったところ、両テストにおける一定時間当たり語彙数の平均には有意差があった($p<.05$)。また、標準偏差はグループテストでの値がインタビューテストのものより大きかったため、グループテストでの受験者の発話量にばらつきが大きいことが示唆された。

表6 発話量についての記述統計及びt検定の結果

	インタビュー		グループ		t検定	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>p</i>
発話量	59	7.1	50	9.7	-3.76	.02

(注) $N=54$, $df=53$

5. 考 察

本研究では、インタビューとグループのテスト形式の違いが受験者発話に及ぼす影響を調べた。検定の結果に有意な違いが見られた言語の特徴としては、流暢さの2指標と発話量であった。次に、流暢さ、正確さ、複雑さ、発話量の言語的特徴について、上述のような結果になった理由をそれぞれ考察していく。

第一に、流暢さの2指標「有声ポーズ数」、「無声ポーズ数」について、統計的に有意な違いが見られたことから、グループテストでの発話がインタビューテストでの発話よりポーズの数が少なく、流暢だったことが示唆された。試験後に受験者たちの感想を聞いたところ、「グループテストでは、他の人が話している時、自分の番の準備ができる」といったような感想が述べられていた。上記に類する回答以外にも、グループテストでは被験者たちに会話維持・管理に自由な裁量を与えられたため、受験者自身が自分なりに言えることを選択できるという点も、流暢に話せたことに繋がったと考えられる。対照的に、インタビューテストでは教師の質問を理解できないと、答えられないため、まず発話以前の段階から教師による言語レベルの高い発話理解と、それに基づく考えの整理が求められるため、言語情報処理に負担が掛かった可能性がある。

この点について、曲(2009)はスピーキングテストの形式の違いが受験者の得点に与える影響を分析した際にも、「流暢さ」の得点で影響を確認した。グループテストで受験者が流暢に話せる理由として、今回の受験者の回答から得られた「自由に発話できた」という情報提供を支持するのであれば、グループ

形式で特定の文法・語彙を見るためのテストとしては不向きな側面があるかもしれない。しかし、会話の維持・管理能力、及び順番取り（turn-taking）など言語の機能的な側面を評価するのであれば十分に活用の価値があるのではないと思われる。その場合、新たな言語教育の配慮や評価項目の作成が必要になってくるであろう。

第二に、正確さについては、2つの指標とも統計的な有意さが認められなかった。つまり、両テストの形式違いが受験者の正確さに影響しないことが示された。本来ならば堀川（2006）の指摘にあった通り、会話内で相手の使った語彙や文法的表現を取り入れて話すことができるため、中国語母語話者の教師が対話者であるインタビュー形式の方がより正確な会話になると当初から予測していた。しかし、今回の結果では両指標ともグループ形式の方に若干高い平均値を示すのみで、その差も有意までには至らなかった。その理由として可能性があるのは、受験者の要因と、参加形態によって取られる方略の違いによる2点が考えられる。まず受験者要因について詳述すると、インタビューテストで正確さが向上しなかった理由として、今回の受験者が高校生であったためと考えられる、つまり、まだ中国語の言語情報処理能力が未成熟な初・中級学習者にとって、インタビュアーの質問に適切な応答をする一連の活動には、言語処理が多岐に及んでいた可能性があった。結果的に相手が使った表現を自分の発話に取り入れるまでには情報処理能力に余裕がなかったと推測できる。この点について、言語能力との相互作用の視点を取り入れ、今後の研究を更に展開していきたいと考える。また、グループでの発話がインタビューテストと同等に正確だった理由として、方略的見地から述べると、グループテストでは、他の受験者が発話している間に自分が何を言うか準備できたことが挙げられる。結果的に、その隠れた準備時間の確保こそが受験者の発話の正確さに繋がったと考えられる。

第三に、複雑さに関しても、2つの指標からは有意な違いが見られなかった。一般的に、複雑な発話が見られるかどうかについては、テストの話題、受

験者の言語レベルにも関係する。平素から使用している教科書から抜粋したタスクに一因があったのかもしれない。また、今後異なる言語レベルの受験者を対象に異なる話題で再検討する必要もあると思われる。もうひとつ考えられる理由として、今回のような利害関係の高いテスト場面では、なるべく減点されずに、正確に言おうと受験者たちが企図していたと考えられ、誤用を恐れずに様々な統語形式にチャレンジする意欲が低かった可能性もある。従って、複雑さに関する結果はテスト形式の違いだけでなく、評価からもたらされる決定事項の重要性が積極的な言語使用への抑制にもなったと思われる。

最後に、一定時間における受験者の発話量が、グループ形式の方で少なかったことについて考察する。通常、言語熟達者であるインタビュアーと会話するインタビューテストでは、テストに対する不安が高まると考えられている。Fulcher (1996) もインタビューよりグループテストの方で受験者のテスト不安が低いと述べているが、今回の発話量の分析結果からは、逆にインタビューの方が安心感を示す受験者が多い可能性を示唆していた。発話量に影響する要因には、テストの形式以外にも、受験者の言語能力、性格、そしてインタビュアーの要因も発話量に多かれ少なかれ影響すると考えられている。(Kobayashi, Johnson, & Van Moere, 2005)。とりわけ今回のインタビュアーは普段の科目担当の教師であったため、テスト不安は極めて低かった可能性がある。また、グループ形式での発話量のばらつきが大きかったことから、グループテストには、たくさん話す者とあまり話さない者の広い格差のあることが分かった。発話量が評価に影響すると考えるならば、より多くの受験者が一定した量の発話を引き出せるためにテスト改善が必要であると思われる。

6. 本研究の意義と今後の課題

本研究では、高校生の中国語学習者を対象にグループとインタビューのテスト形式違いが受験者たちの発話に与える影響について分析を行った。結果、両テストの形式の違いは発話の流暢さ、発話量に影響し、正確さ、複雑さに影響

しないことが明らかとなった。外国語教育現場の大規模教室においては、グループテストをスピーキング能力の評価に用いることによって、時間的なコスト削減ができるばかりではなく、学習者のスピーキングへの学習、教室内でのペア活動、グループ活動にもよい波及効果が期待できる。また、学んでいる外国語を使う機会を与えることで、更なる学習への動機にも繋がるであろう。

従来のスピーキングテストの発話分析は、小規模の調査がほとんどであったが、本研究はより一般化できるように、より多くの受験者を基盤としたデータ収集、及び分析に努力した。また、英語学習者を対象にした研究が多くを占める一方、本研究は中国語学習者を対象とした点でその特質の多くを明らかにしてきた。

しかし、以下に述べる統制の課題もあり、については更なる調査方法の熟考と実証が求められる。とりわけ、テスト形式の影響は、受験者の言語能力、性格、テストの組み合わせ、学習動機などとの相互作用を伴うものである。今後、こうした相互作用へのアプローチも視野に入れ、卓越した調査方法の発展と検証が期待されよう。本研究が、その良き第一歩となれば幸いである。

注

- 1) 使用教科書：相原茂著（2005）『新概念中国語会話—訊くが勝ち』朝日出版社
- 2) AS ユニットは「従属節を含む、1つの独立節または副節単位からなる一人の話者の発話」と定義される（Foster, Tonkyn & Wigglesworth, 2000）。
- 3) Foster（2000）の定義を参考に、計測を行った。具体的に以下の基準で分析を行った。
 - ・節とは主語と独自の時制を持つ定形動詞の組み合わせを持つ語の集まりをいう。
 - ・独立節（an independent clause）：動詞を含む最小限の節
例：| 我吃饭（私はご飯を食べる）| | 他买手表（彼は時計を買う）|
 - ・独立の副節単位（an independent sub-clausal unit）：
 - ・1つかそれ以上の句に対して、談話や場面の文脈から省略された要素が復元可能で、完全な節を構成しうるもの。
例1：| 他请我吃饭。| （彼は私を御馳走する）
例2：| 你来日本几年了？（日本に来て何年？）| | 两年（二年）|

・マイナーな発話。Irregular sentences や Nonsentences と分類されたもの。

例：| 谢谢（ありがとう）| | 好（OK）|

参 考 文 献

- 曲明 2009.「中国語グループ・オーラルテストにおけるテスト得点の分析」,『中国語教育』7:150-169, 中国語教育学会。
- 堀川有美 2007.「日本語会話テストにおいてテスト形式が受験者の発話に与える影響」,『人間文化論叢』第9巻, 255-264 お茶の水女子大学大学院人間文化研究科。
- Bonk, W. J. & Ockey, G. J. 2003. A many-facet Rasch analysis of the second language group oral discussion task. *Language Testing*, 20 (1), 89-110.
- Ellis, R. 2003. Task-based language learning and teaching. Oxford University Press.
- Ellis, R. & Yuan, F. 2004. The effects of planning on fluency, complexity, and accuracy in second Language writing. *Studies in Second Language Acquisition*, 26, 59-84.
- Fulcher, G. 1996. Testing tasks: issues in task design and the group oral. *Language Testing*, 13 (1), 23-51.
- Fulcher, G. 2003. Testing Language Speaking. Edinburgh: Pearson Education
- Foster, P. & Skehan, P. 1996. The influence of planning and task type on second Language performance. *Studies in Second Language Acquisition*. 18, 299-323.
- He, L. & Dai, Y. 2006. A corpus-based investigation into the validity of the CET-SET group discussion. *Language Testing*, 23 (3), 370-401.
- Hiltsdon, J. 1991 The group oral exam: advantages and limitations. *Language testing in the 1990s: communicative Legacy*. 189-197.
- Kobayashi, M., Johnson, K. & Van Moere, A. 2005. Effects of Quantity and Quality of Students' output in Group Oral Tests. 神田外語大学言語教育研究所『言語教育研究』16号, 275-295.
- Luoma, S. 2004. Assessing Speaking. Cambridge: Cambridge University Press.
- Lazaraton, A. 1992. The structural organization of a language interview: a conversation analytic perspective. *System* 20, 373-386.
- Morrison, D. M., & Lee, N. 1985. Simulating an Academic Tutorial: A test Validation Study, 85-92. *New directions in language testing*.
- McNamara, T. F. 1996. Measuring second language performance. Longman
- Mehnert, U. 1998. The effects of different length of time for planning on second language Performance. *Studies in Second Language Acquisition*, 20, 83-108.
- Nagasawa, K., & Tanabe, K. 2001. Interactive English Forum 1999-Trials for improving practical communication skills in Ibaraki prefecture. Ibaraki University, the Faculty of Education.
- Robinson, P. 2001. Task complexity, task difficulty, and task production: Exploring interactions in

- a componential framework. *Applied Linguistics*, 22, 22-57.
- Richards, J. C, 2001. Curriculum development in language teaching. Cambridge University Press.
- Swain, M. 2001. Examining dialogue : another approach to content specification and to validating inferences drawn from test scores. *Language Testing*, 18 (3), 275-302.
- Taylor, L B 2001. 'Study of Quantitative Differences between CPE Individual and Paired Speaking Tests, ' Internal UCLES EFL Report.
- Young, R. & Milanovic, M. 1992. Discourse variation in oral proficiency interviews. *Studies in Second Language Acquisition*, 14, 403-424.
- Van Moere, A. 2006. Validity evidence in a university group oral test. *Language Testing*, 23 (4), 411-440.
- Van Lier, L. 1989. Reeling, Witching, Drawling, Stretching, and Fainting in Coils : Oral Proficiency Interview as Conversation. *TESOL Quarterly*. 23 (3), 489-508

付録1. グループテスト用タスク

- ① 中国人の高校生が東京に一日だけ滞在することになりました。その日に、東京観光とお土産を購入する予定です。東京の面白いところはどこですか。その理由は。お買い物はどこのだんな店に行った方がいいと思いますか。どうしてでしょうか。また、その店に行くのにどのような交通手段を用い、どのぐらい時間が掛かりますか。どうぞみんなでアドバイスをしてあげてください。
- ② 来週中国の高校生との交流会が控えている。中国語を用いて、自分たちのこと、そして日本の高校（特にXX学院）の日常生活を紹介することになりました。三人でその紹介文を作ってください。具体的に、自分たちの名前、身分、趣味、家族のこと、中国語の学習歴、それから学院の生活、何時に起きる、寝る、何時に授業開始、終わり、部活などのことを紹介するようにしてください。
- ③ 三人で学園祭のお芝居の主人公を作りあげることになりました。どのような主人公を作り上げるのかについて、話し合ってください。具体的に、主人公の外見、様子(顔、目、髪の毛、体つき、着ている洋服、持ち物)、趣味、家族のことなどについて決めてください。

付録2. インタビューテストで用いる質問文例

- ・你学汉语学了多长时间了？喜欢汉语吗？为什么？
- ・你是大学生吗？几年级？一天几节课？什么课？业余时间喜欢做什么？
- ・你几点（什么时候）开始上课（放学，起床，睡觉，放寒假，放暑假等）？
- ・你们的汉语老师（爸爸，妈妈，哥哥，朋友等）长得什么样？穿着什么样的衣服？喜欢什么？
- ・这手表（衣服，眼镜等）是在哪里买的？去学校（医院，商店等）怎么走（去）？

- ・ 你给爸爸, (妈妈, 哥哥, 朋友等) 买了什么?
- ・ 东京哪里好玩儿? 为什么? 你经常在哪里玩?

Summary

This paper investigated the effects of the test format on the students' performance in a Chinese speaking test by analyzing students' discourses during the following two tests. A total of 54 students, took two types tests: the group format and the interview format as their term examination of Chinese Speaking course. Discourses were analyzed in terms of fluency, accuracy, complexity and the amount of learners' output. The results showed students spoke more fluently in a group test than in an interview test, but the amount of the output is more in the interview than in the group test.

Through the research on the effects of test formats, the group oral yields valid estimates of a test taker's oral ability can be identified, and optimal design principles for the test format can be ascertained.

Keywords: oral test, group format, interview format, test performance

本稿は, 松山大学特別研究助成(平成21年度)による研究成果の一部である。